Deepfake Article Review

Keyon Gordon

Old Dominion university

CYSE-201S - Cybersecurity and the Social Sciences

Diwakar Yalpi

9/28/25

Relation to Social Science Principles

The article connects AI with social science by presenting how deepfakes are used to trick the human mind and perception into believing what they're seeing is real. One example of a core social science principle that was used to conduct this study is empirical data. The use of online surveys and experiments where each participant was given an image and had to label it "deepfake" or "real" is a prime example of empirical data being taken (Bray, 2023). This article showed how you can include social science principles to understand how humans interact with AI/deepfake content.

Research question, Hypothesis, IV/DV

From reading the article, some of the research that was brought up is "How well can humans identify faces generated by AI?" and "Does a participant's confidence relate to accuracy?" (Bray, 2023). These were the main questions that the experiment was built and conducted around. Unfortunately, the article never states a formal hypothesis. The independent variable for this experiment was the images used. Some images were easier than others, according to the participants. The dependent variable was the accuracy with which the participants detected the deepfakes and the confidence in their decision-making. The main takeaway from this experiment was whether the image used influences the accuracy and confidence of the participant.

Research Methods

In this experiment, each participant was given 20 images with a mix of real and deepfake images within. The participants had to detect which image was a deepfake and how confident they were in their answer. They were also given tell-tale signs of when an image is a deepfake (Bray, 2023).

Types of data and analysis

Quantitative data was used to measure the participants' answers. Binary correctness was a big method used to measure the participants, whether an image was correctly labelled or not. Their confidence in their answers was also measured on a numeric scale, further showing how big a role quantitative data played in this experiment.

Relation to PowerPoint slide

This week's PowerPoint slide focused on cybercrimes and victims of cybercrime. Deepfake images coincide with cybercrime. Over the years, AI has gotten more advanced, making the images generated much more realistic than before. We have seen celebrities speak out about

deepfake images of their face being used inappropriately without their consent. Lawsuits have been filed attacking people who make these deepfakes, and laws have been passed to restrict their use of deepfakes.

Relation to Challenges, Concerns, and Contributions of Marginalized Groups

A concern that I have for marginalized groups is older people and people with media literacy. Every day, deepfake images get more realistic, and older people and those who aren't familiar with social media can easily fall prey to these images, believing that they are real. We have also seen political figures fall for deepfake images. Although this article doesn't specifically study marginalized groups, history proves that they are vulnerable.

Overall contribution

This study magnified how easily humans are tricked by deepfake images. This shows that assistance or training needs to be put in place regarding deepfake images, showing different techniques to detect them. This study was also great for connecting human judgment with AI. Using AI-generated images to test the judgments of the participants encourages the use of more AI/human studies being conducted.

Conclusion

This study perfectly combined the use of social science with computer science. Using humans to judge whether images are real uses human perception to make a correct judgment. While combining the use of social and computer science, it also magnified how hard it can be for people to identify AI-generated images, and with the advancement of AI, it will only get harder. Making the need for technical solutions to be created and pushed appropriately.

Reference

Bray, S. D., Johnson, S. D., & Kleinberg, B. (2023). Testing human ability to detect "deepfake" images of human faces. *Journal of Cybersecurity*, *9*(1). https://doi.org/10.1093/cybsec/tyad011